The Importance of Biological and Technical Controls in the Application of Methylation-Sensitive Amplified Fragment Length Polymorphism

Tyler W. Cunningham, Brendan G. Hunt, Soojin Yi, and Michael A.D. Goodisman

Georgia Institute of Technology

School of Biology

310 Ferst Drive, Atlanta, GA 30332

Abstract

DNA methylation is an epigenetic modification that is involved in a wide array of biological processes, including disease and aging. Methylation research involving insects has been limited in the past due to the lack of complete methylation machinery in model organisms like *Drosophila melanogaster*. However, a number of insects with complete sets of DNA methylation enzymes have now had their genomes sequenced, but there is still great interest in methods that assess global DNA methylation levels without genome resources. In this study, the efficacy of methylation-sensitive AFLP is investigated. Specifically, we tested this protocol's sensitivity to methylation differences and replicability. We attempted to compare methylation levels of many different organisms, such as *Vespula maculifrons, Apis mellifera, Caenorhabditis elegans*, *Homo sapiens,* and *Drosophila melanogaster.* In our experiments, we observed poor reproducibility between replicates. Furthermore, *D. melanogaster* and *C. elegans*, which do not exhibit significant levels of DNA methylation, exhibited particularly inconsistent and spurious results. In this study, we systematically tested each step of the methylation-sensitive AFLP protocol to identify possible sources of error. We find that, while most of the variables of the methodology are robust, some organisms and the methylation-sensitive enzyme, HpaII, do not return consistently reproducible results. Therefore, if one chooses to implement this methodology, biological and technical controls must first be proven to be working as expected before any "true" comparative studies may be conducted.

INTRODUCTION

DNA methylation and its role in organisms

The DNA of many organisms is modified through methylation. Specifically in animals, cytosine residues followed by a guanine, otherwise known as CpG dinucleotides, are sometimes methylated by enzymes called DNA methyltransferases (DNMTs). DNA methylation of these cytosine nucleotides is an important and prevailing epigenetic process in many eukaryotes (Han *et al.*, 1999).

DNA methylation appears to have a wide array of effects in eukaryotic organisms. In humans, cancer may be associated with epigenetic changes, such as hypermethylation of certain DNA sites (Brena *et al.*, 2006). Furthermore, global DNA methylation levels differ according to gender and decrease as humans and other organisms' age (Boks *et al.*, 2009; Fuke *et al.*, 2004; Kronforst *et al.*, 2008). The process of DNA methylation has also been implicated in genomic imprinting (Boks *et al.*, 2009; Kronforst *et al.*, 2008). From all of these and other examples, we have learned that DNA methylation has diverse functional significance in a wide range of organisms.

DNA methylation is a highly conserved phenomenon found throughout different species (Johnston *et al.*, 2005; Rozhon *et al.*, 2008; Suzuki, Bird, 2008). Yet there is relatively little known about the role of genomic methylation in insects due to the conflicting information present in the literature and the lack of a suitable model organism in which to study DNA methylation (Wang *et al.*, 2006). Specifically, DNA methylation is virtually absent in the model insect *Drosophila melanogaster* (Field *et al.*, 2004; Marhold *et al.*, 2004). This finding originally

led researchers to believe that insect genomes would not contain DNA methylation. Recent studies, however, have shown that some insects do contain the molecular machinery (i.e. DNA methyltransferases), necessary for DNA methylation. In particular, the genome of the honey bee, *Apis mellifera,* has been sequenced and been shown to possess discernable levels of DNA methylation, which has opened the door for methylation studies in insects, including parasitoid wasps and pea aphids (Kucharski *et al.*, 2008; Walsh *et al.*, 2010; Wang *et al.*, 2006; Werren *et al.*, 2010).

The role of DNA methylation in social insects

Kucharski et al. (2008) showed that DNA methylation of CpG dinucleotides plays a role in caste differentiation (i.e. the differentiation of a larvae to the worker or queen phenotype) within *A. mellifera*. Utilizing RNAi techniques, Kucharski et al. were able to down regulate the enzyme DNMT3, a methyltransferase, in pupae. This down regulation caused pupae that were originally destined to express the worker caste phenotype to show the queen phenotype. This result suggested the DNA methylation might play an important role in modulating social phenotypes and behavior in insects. Thus, we hypothesized that global DNA methylation plays a crucial role across the social insect caste.

The original motivation for this study was to understand and determine levels of DNA methylation in social insects other than *A. mellifera*. In particular, we were interested in determining if the social wasp, *Vespula maculifrons*, displayed discernable levels of DNA methylation. Results from earlier studies led us to expect that the levels of global DNA methylation differ between organisms of the same species, and between species, such as *A. mellifera*, *D. melanogaster*, and *V. maculifrons* (Kronforst *et al.*, 2008; Kucharski *et al.*, 2008).

To analyze the global levels of methylation between organisms, we required a method that would allow us to analyze many individuals in a fast, inexpensive, and accurate way.

Global Methylation Assessment Methods

A multitude of molecular techniques are available for determining global levels of DNA methylation. For example, levels of methylation within the genome can be determined using high performance liquid chromatography (HPLC). HPLC requires only a low amount of starting DNA to determine the presence of the nucleotides adenine, cytosine, thymine, guanine and methylated cytosine. The method works by eluting off compounds based on their molecular weights (Fraga, Esteller, 2002). This particular methodology is applicable to assessing methylation patterns because all the nucleotides, including methylated cytosines, have different molecular weights and therefore elute off at different times during the experiment.

A second method of determining levels of DNA methylation is bisulfite sequencing, which has become known as the "gold standard" in methylation studies. This method converts epigenetic differences into nucleotide substitutions (Laird, 2010). In bisulfite conversion, unmethylated cytosines are transformed into thymines. The target DNA region can then be sequenced to determine the genetic code of the sequence in question, and analyzed at the CpG sites for a change from a cytosine in the pre-bisulfite digestion to a thymine in the post-bisulfite digestion (Laird, 2010;Walsh *et al.*, 2010). One can then assume that any cytosines present in the sequence must be methylated and were protected from the bisulfite digestion.

Another useful technique for global methylation studies is methyl-CpG immunopreciptation (MCIp) (Schilling, Rehli, 2007). In this method, a fluorescent antibody is created that attaches to methyl-cytosines. The labeled product is then hybridized to a micro-array

for analysis. Difficulties with this method arise from variation in the experimental protocol that must be made for different tissues and organisms due to the different levels of CpG dinucleotides in their DNA. Yet, the method's power comes from quick whole genome assessments of an organism's level of methylation (Laird, 2010).

The last molecular technique of interest, Methylation-Sensitive Amplified Length Polymorphism (AFLP), uses two independent endonuclease reactions which use methylation-sensitive restriction enzymes and methylation-insensitive enzymes to create an estimate of the proportion of methylated to unmethylated target sites in a genome (Han *et al.*, 1999). Methylation-sensitive AFLP relies on the fact that most methylation in animals occurs at CpG sites by using enzymes that target and preferentially amplify these sequences in the genome. The methylation insensitive enzyme should always cut at these CpG sites and the methylation-insensitive enzyme activity should be blocked if a CpG site is methylated. These two reactions, when visualized, result in two comparable profiles which should differ almost exclusively at methylated sites. We used the AFLP methodology because of its ability to return highly reproducible results, generate large amounts of data from a single assay, and has the ability to use this method without sequence data or specific predictions (Trybush *et al.*, 2006). However, we suggest that methylation-sensitive AFLP requires extensive validation with biological controls which is frequently overlooked (Han *et al.*, 1999).

Our original motivation in this study was to determine global levels of methylation in the social wasp, *V. maculifrons*. We first decided to test if methylation-sensitive AFLPs successfully displayed actual levels of DNA methylation, by analyzing global levels of methylation in three control species, *Homo sapiens*, *A. mellifera*, *D. melanogaster, and C. elegans*. Three of these species, *D. melanogaster, H. sapiens*, and *C. elegans* have well characterized values for their

overall level of global DNA methylation at CpG sites (Field *et al.*, 2004; Fuke *et al.*, 2004). *D. melanogaster* and *C. elegans* display very low levels of methylation and were used as our 'negative controls'. In the context of our system, a negative control refers to the complete absence of DNA methylation or its presence at very low levels. These organisms should exhibit few polymorphisms between the profiles of methylation-sensitive and methylation-insensitive enzymes. *H. sapiens* possesses a high level of DNA methylation at CpG sites and was used as our 'positive control.' A positive control is expected to exhibit many polymorphisms between the profiles of methylation-sensitive and methylation-insensitive enzymes. The final species in our initial tests, *A. mellifera,* was used because its level of methylation is somewhere between that of our positive and negative control species (Wang *et al.*, 2006). Thus, we predicted that the measurable magnitude of global DNA methylation would decrease in the following order: *H. sapiens*, *A. mellifera*, and *D. melanogaster/C. elegans*. We hypothesized that *V. maculifrons* would have a methylation profile similar to that of *A. mellifera* because they are members of the same insect order.

This report describes our attempts to obtain reliable results using the methylation-sensitive AFLP methodology. Based on our difficulties producing reliable, biologically relevant results, we suggest that extreme care be taken when using this technique. Here we describe sources of variation that may arise from methylation-sensitive AFLP and suggest experimental approaches for validating this technique.

MATERIALS AND METHODS

Our original methodology was adopted from the protocol of Kronforst et al. (2008). Genomic DNA (~100 ng/μL) from selected organisms was obtained using a Qiagen (Valencia, California) DNeasy tissue kit. Two aliquots of the same DNA sample were each subjected to different enzyme restriction reactions, an EcoRI-MspI reaction and a EcoRI-HpaII reaction. The reaction protocol is as follows for the EcoRI-MspI restriction: 3μL of the target DNA, 0.05μL EcoRI (5U), 0.25μL MspI (5U), 1μL NEBuffer 2, 5.7μL dH_20 (all reagents obtained from New England Biolabs). The EcoRI-HpaII restriction reaction protocol is as follows; 3μL of the target DNA, 0.05μL EcoRI (5U), 0.5μL HpaII (5U), 1μL NEBuffer 1, 5.45 μL dH_20 (all reagents obtained from New England Biolabs). EcoRI recognizes the sequence 5'-GAATTC-3', while MspI and HpaII recognize the sequence 5'-CCGG-3'.

The key to this methodology is that the EcoRI-MspI digestion and the EcoRI-HpaII digestion are differentially affected by cytosine methylation. The MspI digestion's activity is not affected by DNA methylation and the HpaII digestion is blocked by CpG methylation (Kronforst *et al.*, 2008). These two reactions, EcoRI-MspI and EcoRI-HpaII, were incubated at 37°C for three hours.

The products of both of these digestions were then ligated with EcoRI and MspI-HpaII adapters. The adapter ligation reaction was performed using 3μL of the restriction product, 1μL of the EcoRI adapter (5pmol; Table 2), 1μL of the MspI-HpaII adapter (50pmol; Table 2), 0.25μL T4 DNA ligase (400,000U/mL), 1μL ligase buffer (New England Biolabs), and 3.75μL of dH_20. This reaction was performed on both the MspI and HpaII restriction at 37°C for three hours and then left at room temperature overnight. The ligation products were then diluted with 100μL of dH_20.

The two reactions were subjected to two rounds of PCR to attain maximum amplification of the desired products. The first PCR, hereafter referred to as the pre-select PCR, was conducted to amplify the ligation products from the previous step using a one nucleotide selective extension. Therefore, two unlabeled primers MHpre and EcoRIpre (1μL each) (Table 2), core mix (7μL) (Table 1), and ligation product (1μL) were used to achieve this goal. The thermocycler profile for this step is as follows: twenty cycles of the following pattern; 94°C for thirty seconds, one minute at 60°C, and one minute at 72°C. PCR products were then diluted with 100μL of dH_20.

Table 1. Table showing the volume of reagents used in creating a core mix for this protocol. Core mix used in non-select and select PCR.

Core Mix:	1x (μL)
Sterile ddH_20	4.10
25mM $MgCl_2$	0.80
10X PCR Buffer	1.00
dNTPs (2.5 mM)	0.80
Taq (~5U/μL)	0.30
Total Volume (μL)	7.00

The second PCR, the select PCR, selectively amplifies sequences of DNA with a specific nucleotide pattern (Table 2), which amplifies a subset of the available segments of DNA available. These select primers vary in their selective nucleotides by two to six nucleotides. The select PCR reaction consists of two primers, the MH primer, an unlabeled primer (1μL), and the EcoRI primer, a fluorescently-labeled primer (1μL; see Table 2 for list of primers), core mix (7μL; Table 1), and pre-select PCR product (1μL). The thermocycler pattern for this round of PCR has the following profile: 36 cycles, 30 seconds at 94°C, 30 seconds at the annealing temperature, and one minute at 72°C. The annealing temperature in the first run was 65°C and was reduced by 0.7°C for the next twelve cycles. The profile continued at 56°C for the rest of the

cycles. PCR products were then diluted with 100μL of dH_20. PCR products obtained through this method were then analyzed on an ABI (Foster City, California) prism 3100 genetic analyzer and quantified using GeneScan software.

Table 2. Ligation adapters, pre-select primers, and select primers used in the experimentation and optimization of the methylation-sensitive AFLP protocol

Adapter/Primer	**Sequence (5' to 3')**
Ligation	
EcoRadpt1	CTCGTAGACTGCGTACC
EcoRadpt2	AATTGGTACGCAGTCTAC
MHadpt1	GACGATGAGTCTAGAA
MHadpt2	CGTTCTAGACTCATC
Pre-Select PCR	
EcoRpre	GACTGCGTACCAATTC
MHpre	GATGAGTCTAGAACGGA
Select PCR	
Eco-AA	GACTGCGTACCAATTCAA
Eco-AC	GACTGCGTACCAATTCAC
Eco-AG	GACTGCGTACCAATTCAG
Eco-AT	GACTGCGTACCAATTCAT
MH-AAT	GATGAGTCTAGAACGGAAT
MH-AATAAC	GATGAGTCTAGAACGGAATAAC
MH-AATAC	GATGAGTCTAGAACGGAATAC
MH-ACT	GATGAGTCTAGAACGGACT
MH-AG	GATGAGTCTAGAACGGAG
MH-ATC	GATGAGTCTAGAACGGATC
MH-ATCA	GATGAGTCTAGAACGGATCA
MH-ATT	GATGAGTCTAGAACGGATT

We attempted to optimize this protocol to attain reliable and accurate results because our preliminary analyses did not fit with predicted differences in the methylation status of our biological controls. In particular, our results indicated that there was a relatively high level of methylation in both of our negative controls, *D. melanogaster* and *C. elegans*. Also, we could find no discernable level of methylation in *A. mellifera*. In the following section, we describe how we varied aspects of the above protocol to determine if variation in particular steps resulted

in variation in the resulting AFLP profiles. Our goal was to determine if our unexpected and nonstandard results were caused by inherent errors in particular steps of the procedure.

Enzyme Restriction

We varied the amount of restriction enzymes used during digestions in order to determine if variation in AFLP profiles occurred because of incomplete digestion. Normal protocol for this step involves genomic DNA at ~100ng/μL and a total digestion time of 3 hours at 37°C. To test our incomplete digestion hypothesis we digested DNA from *M. musculus* for 0 minutes, 5 minutes, 3 hours, and 16 hours. In another experiment, we tested to see if the initial concentration of DNA caused any variation in the profiles. Concentrations at 50 ng/μL, 100 ng/μL, and 200 ng/μL were tested with V. maculifrons DNA. All other aspects of the protocol were as described above.

Ligation Reaction

To determine if the ligation reaction was a source of variation in AFLP profiles, we varied the time in the methodology. The normal setup for this section was completed at 37°C for 3 hours. We experimented with variable incubation temperatures at 16°C, 37°C, and room temperature. Technical replicates with the standard methodology were also produced to determine if the ligation reaction was giving rise to variation in AFLP profiles.

PCR Cocktail

We next investigated if different components of our PCR cocktail could have caused variation in our AFLP profiles. The original setup of our core mix (Table 1) was varied to test if differences in setup would cause noticeable changes in the profiles. In our experiments, the

amount of primer added varied from the control of 5pmol, for each primer, to 10pmol. The Taq polymerase was also changed from the original of 0.3μL to 1μL to see if the increase in enzyme activity would create a more robust profile. Finally, the overall reaction volume was changed from 10μL to 20μL to find if evaporation of the small reaction volume was causing variation between replicates. All reaction variants were tested during the select-PCR step.

Pre-Select PCR

The Pre-Select PCR, the purpose of which is to amplify a subset of restriction products according to a one-base selective nucleotide extension, was removed from the profile in some our experiments to test if this decision would dramatically change the profile or affect the reproducibility of the runs. Technical replicates were also run to test if the step was creating polymorphisms in the profiles.

Select PCR

Throughout experimentation, a multitude of select-primers were used to determine the efficacy and reproducibility of the select PCR step (Table 2). Experiments were run in which primers of different lengths, from three to six bases, were used to see if longer or shorter primers returned more reproducible profiles. It is hypothesized that each extra base should decrease the number of profiles in a given profile by one fourth. As in the above steps, technical replicates were created to test if this step was causing differences in the profiles.

Dilution Protocol

We hypothesized that different dilution protocols could be a source of variation resulting from the decrease in intensity creating miscalls. Therefore, we diluted our select-PCR products with dH_20 in a combination of different ratios; ranging from no dilution to 10:1 dilution of dH_20:product.

Organism Variation

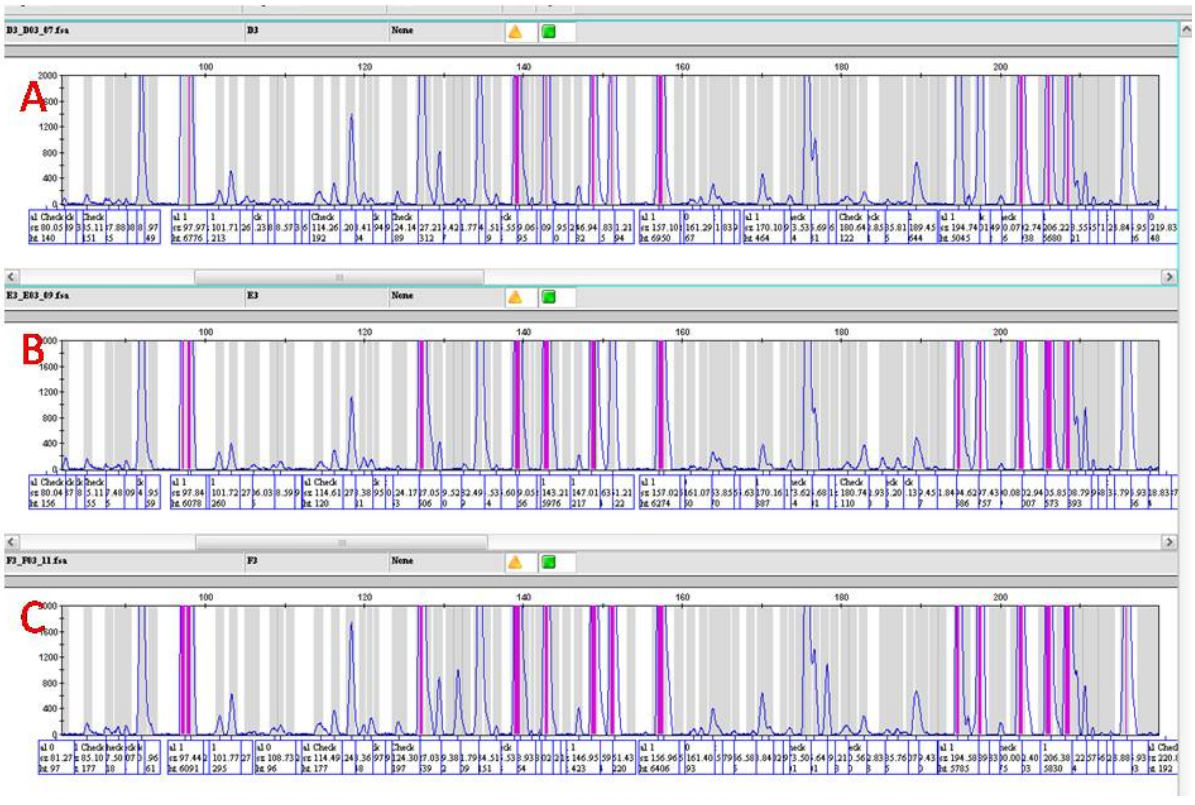
Several different organisms, ranging from vertebrates to invertebrates, were used in the above experiments because we wanted to make certain that our changes would produce the same results in low methylated organisms as it did in highly methylated organisms. To this end, we used *M. musculus* and *H. sapiens* to represent organisms with high levels of methylation organisms. *D. melanogaster* and *C. elegans* were used to represent organisms with low levels of methylation. Finally, *A. mellifera*, *V. maculifrons*, and *Polistes exclamans* were used as our intermediate/unknown group of organisms. These organisms were consistently used throughout our experiments multiple times because of the surprising results that were produced.

RESULTS

We tested various parts of the methylation-sensitive AFLP procedure in order to determine if the protocol was sensitive to these changes. We now describe how changes in each of the tested stages affected our results.

Enzyme Restriction

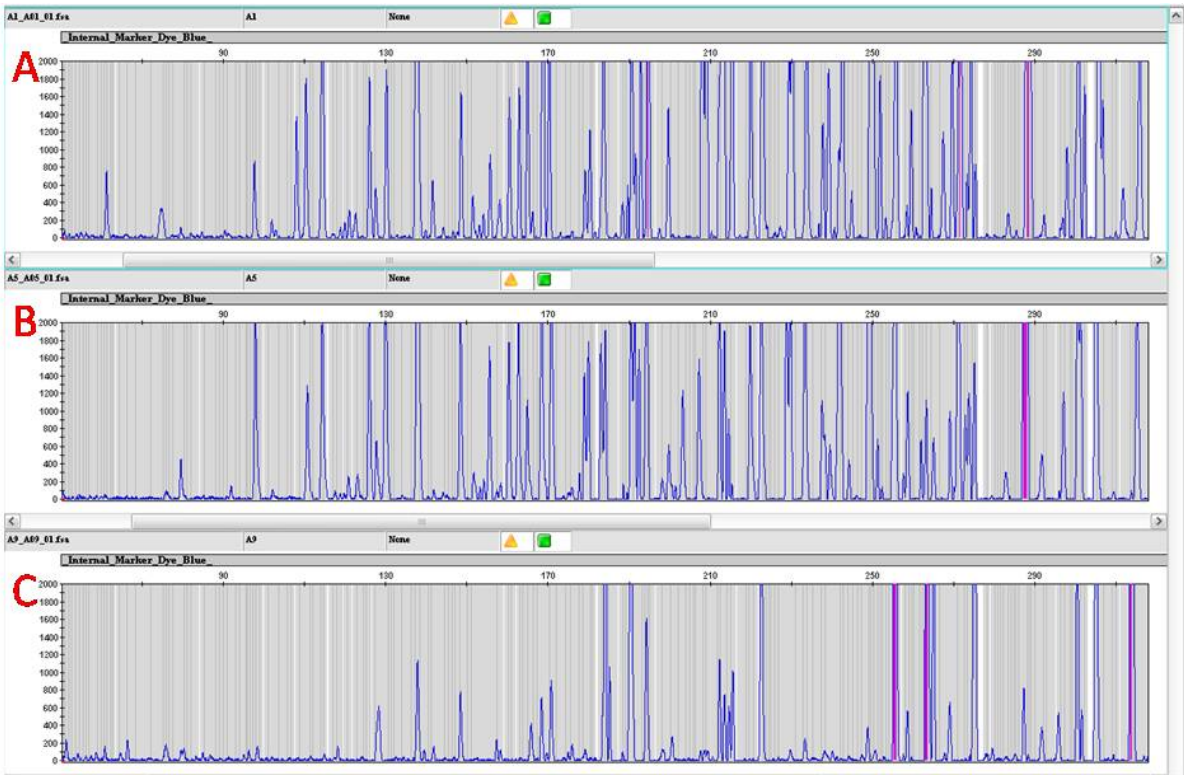
Figure 1. *V. maculifrons* AFLP profile initiated with Eco-AT/MH-ATT with different initial concentration of DNA. Panel A started with 50 ng/µL. Panel B started with a concentration of 100 ng/µL. Panel C started with 200 ng/µL.



The result of the DNA concentration experiment is shown in Fig. 1. As one can see, profiles A, B, and C are all very similar (Fig. 1). This result was consistent throughout all replicates of this experiment (additional profiles not shown). Consequently, we found no

evidence that the amount of genomic DNA template used in the reaction affected the profiles produced.

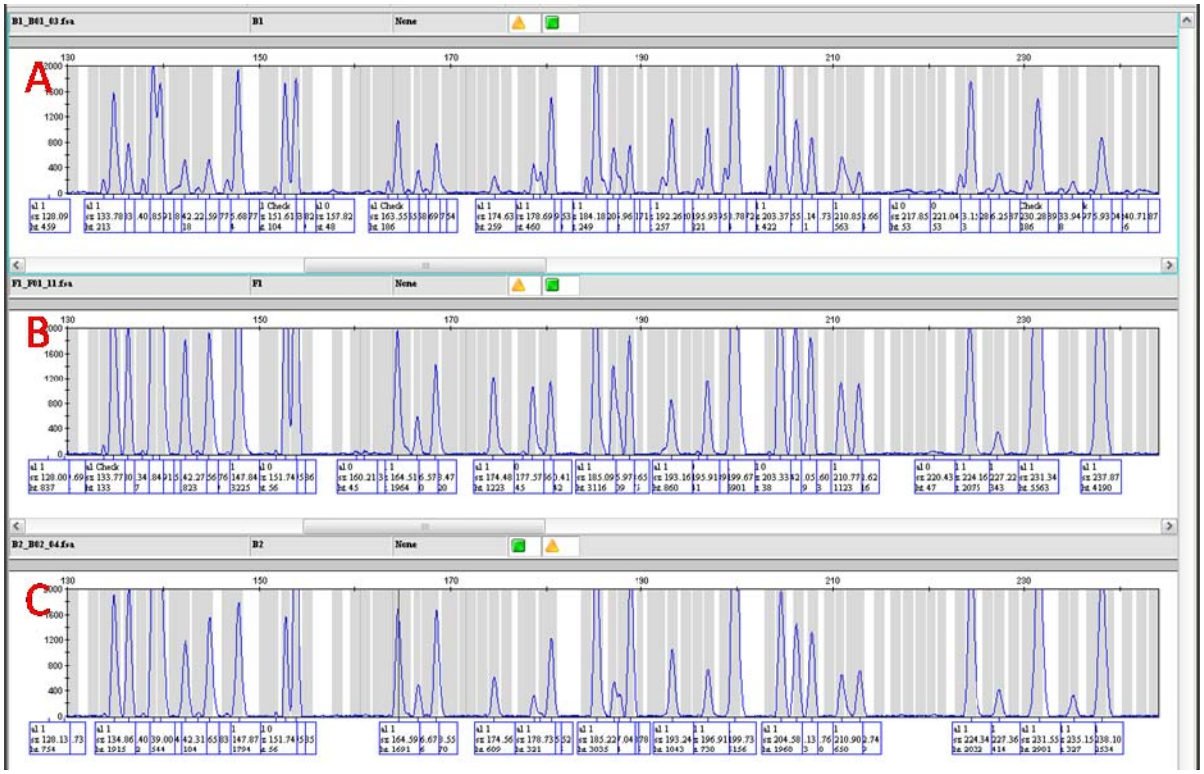
Figure 2. *M. musculus* profile with variable restriction enzyme digestion time. All profiles cut with the MspI enzyme primer, meaning they were initiated with the EcoRI-MspI primer (this shorthand of cut is used throughout this section), set Eco-AT/MH-ATT. A underwent digestion for 5 minutes. B underwent digestion for 3 hours. C underwent digestion for 16 hours. These profiles all underwent the same digestion and underwent the rest of the protocol independently.



We did find an effect of the digestion time on AFLP profiles (Fig 2). We found that digestion of DNA for as little as five minutes produced profiles similar to digestion for as long as three hours (Fig.2A and 2B, respectively). Although, we note that the 5 minute and 3 hour profiles are not identical. However, digestion for 16 hours produced a very different profile with far fewer peaks (Fig.2C). The cause of this could be DNA degradation, although it is unclear from our results. Regardless, we do find evidence that digestion time can influence the AFLP profiles.

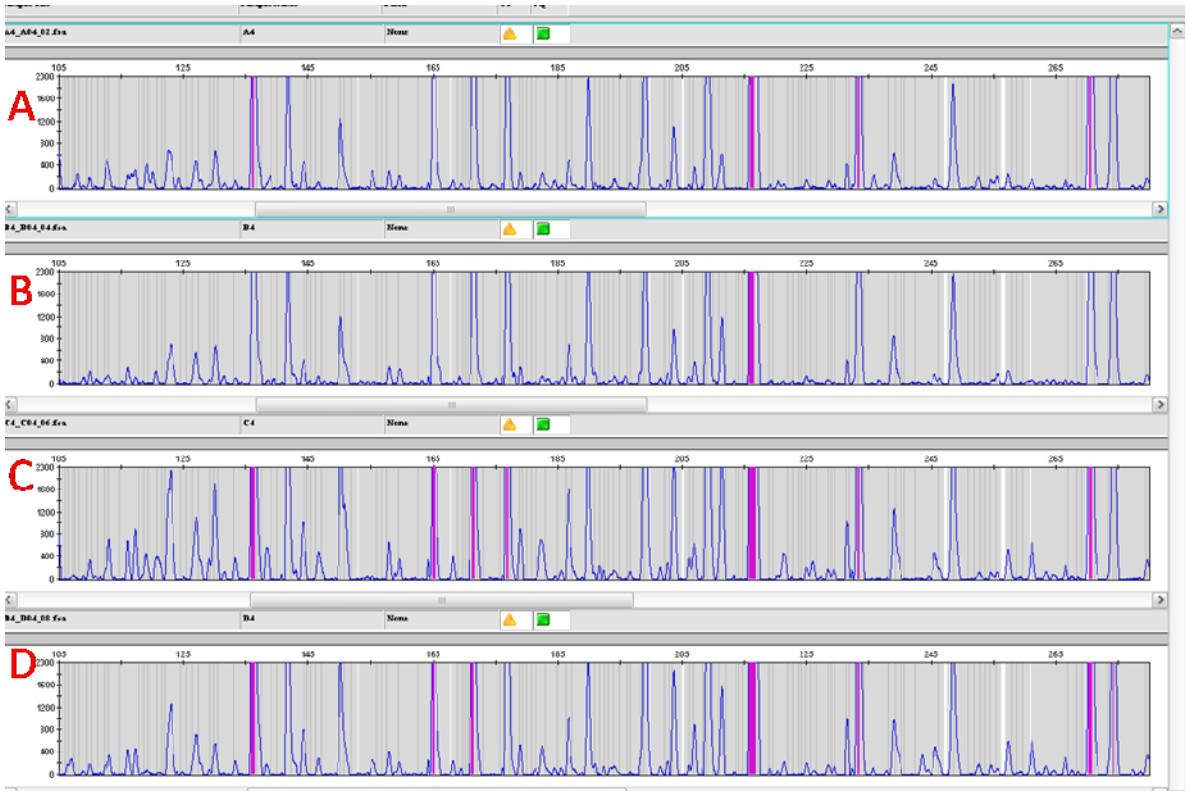
Ligation Reaction

Figure 3. *H. Sapiens* profile with primer set Eco-AC/MH-ACT. Profiles all have different ligation temperatures. A was incubated at 16°C. B was incubated at 37°C. C was incubated at room temperature.



Ligation temperature seems to play some role, though somewhat limited, in the reproducibility of profiles (Fig. 3). We can see that 37°C (Fig. 3B) and room temperature (Fig3C) show general agreement. 16°C (Fig. 3A) deviates from the other two panels in morphology and intensity leading us to make the claim that this is too low a temperature for the reaction to work properly. Ligation temperatures ranging from 25°C to 37°C seem to be appropriate for the ligation adaption reaction to work in a robust manner.

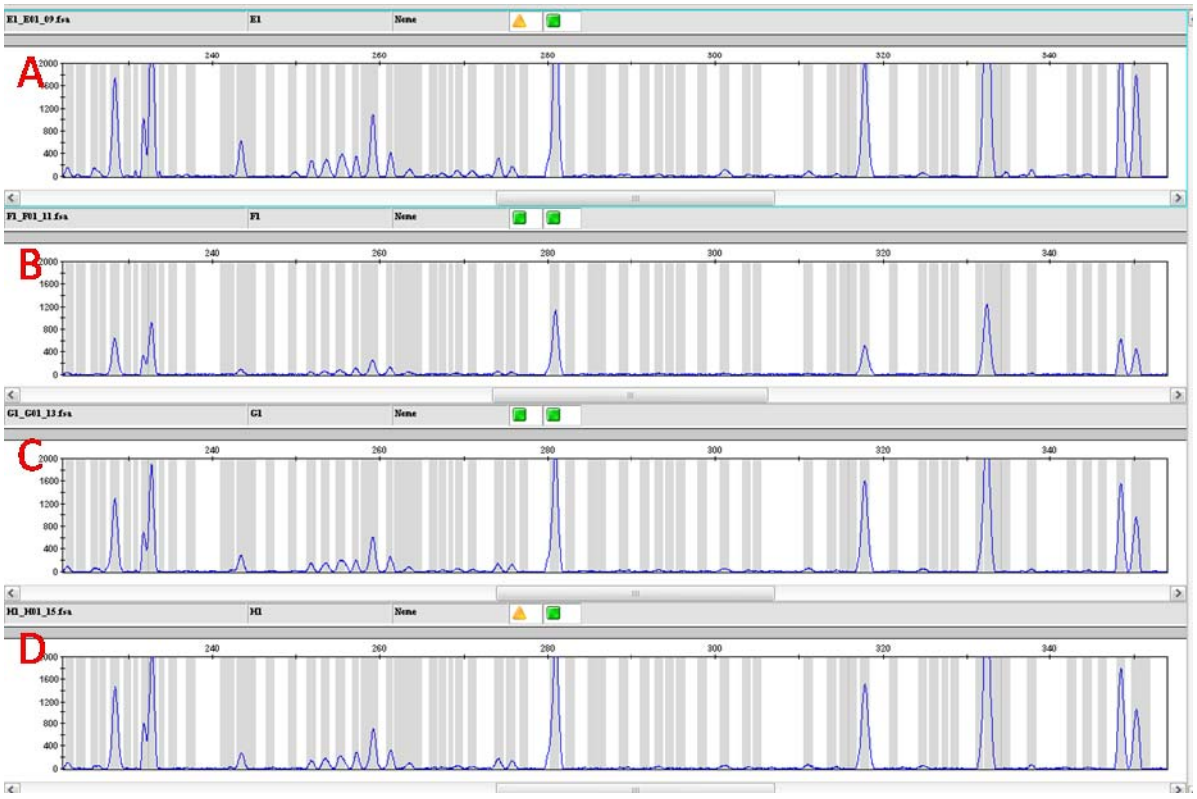
Figure 4. *D. melanogaster* profile cut with HpaII enzyme only, primer set Eco-AT/MH-ACT. A and B are from the same digestion but underwent separate ligation reactions. C and D are from the same digestion, but different from that of A and B, and underwent separate ligation reactions.



In Figure 4, both the reproducibility of the enzyme digestion step and the ligation reaction are under scrutiny. Fig. 4 A and B differ from C and D, but it looks like this might only be due to intensity differences. However, these differences can lead to polymorphisms and it should be noted that all profiles need to be near the same intensity. Comparisons within Fig. 4 C and D and A and B show very high agreement leading us to believe that the ligation reaction is not a source of variability. This result was seen, but to a lesser degree, in the MspI profile (not shown).

PCR Cocktail

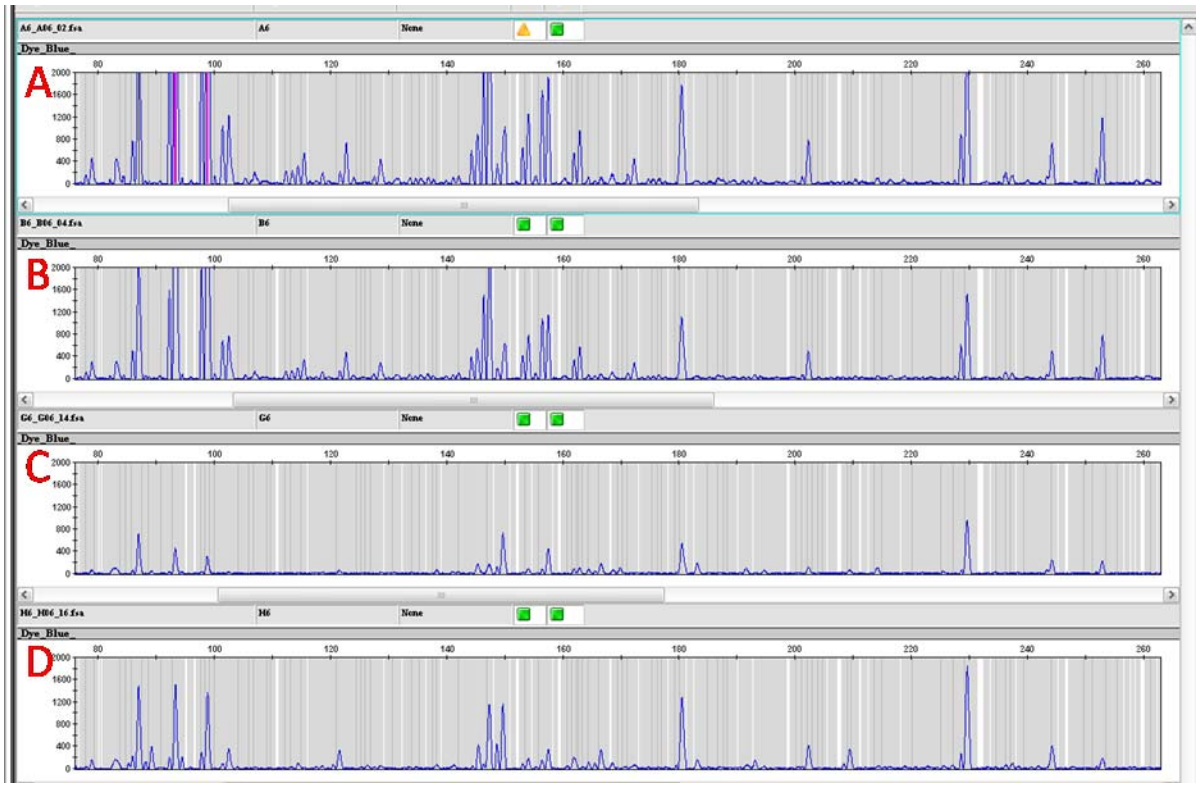
Figure 5. *V. maculifrons* profile cut with HpaII. All panels used primer set Eco-AT/MH-ACT. A is the control with the normal methodology as in the Methods section. B is the normal protocol but with 1µL of Taq. C is the normal protocol with twice the primer added. D is the normal protocol but instead of a 10µL reaction it is a 20µL reaction.



In this section of experiments (Fig. 5) we varied different sections of our PCR cocktail for the select PCR reaction. We can see that changing the protocol does not change the reproducibility of the methods as Fig. 5 A,B,C, and D all look very much alike. The only discernable difference is that increasing the amount of Taq in Fig 5B degrades the profile somewhat. Therefore, we can infer that the PCR cocktail is robust to changes and is not a source of variation within the protocol.

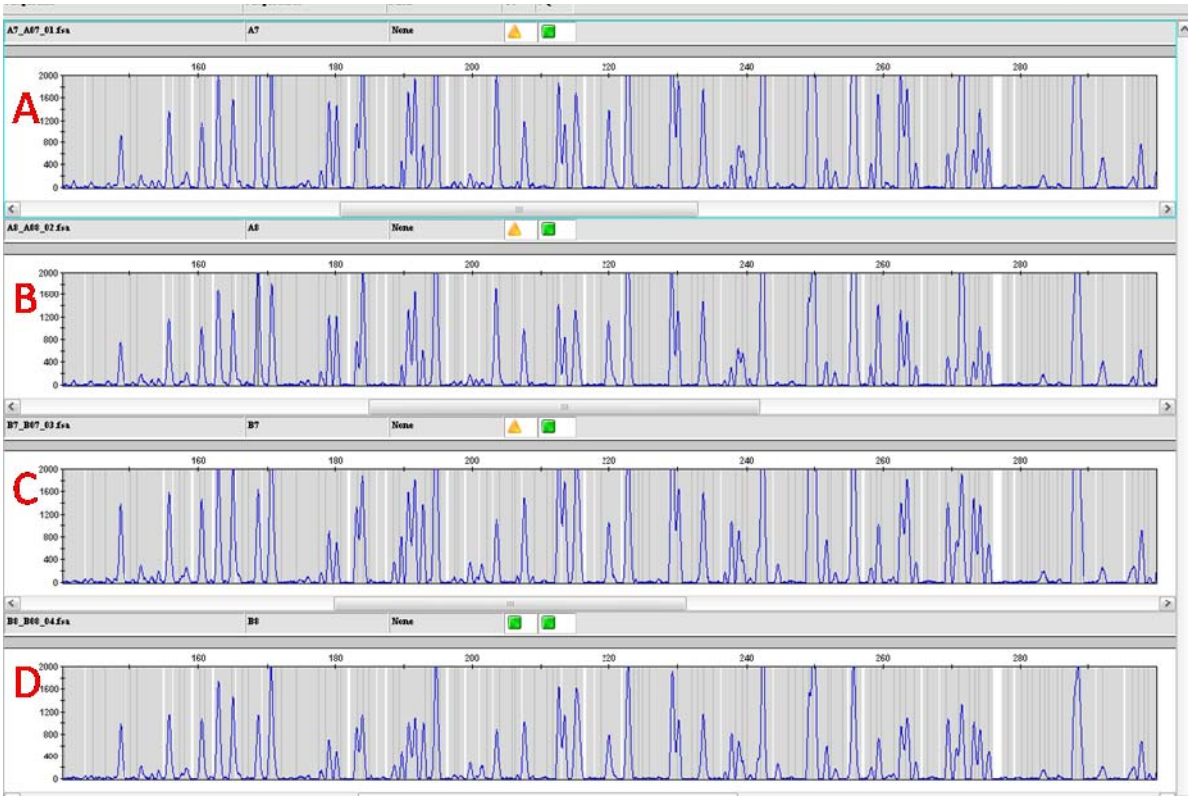
Pre-Select PCR

Figure 6. *H. Sapiens* profile with primer set Eco-AG/MH-ATA. All profiles are of the HpaII enzyme. A and B underwent the same pre-select PCR step and different select PCR steps. C and D did not undergo the pre-select PCR step and independent select PCRs.



With most of our experiments taking a relatively long time to conduct, we experimented with eliminating the pre-select PCR step. This would decrease the overall time for the methodology and would decrease the chance for variation in the protocol. Figs. 6A and 6B show high reproducibility within each other and show good peak morphology during their respective runs. Figs. 6C and 6D show good correlation between each other but their low intensity is a problem. This intensity could create miscalls and indicate polymorphisms where there are none. Therefore, we conclude that the select PCR step should be included for the purpose of increasing the intensity of the profile.

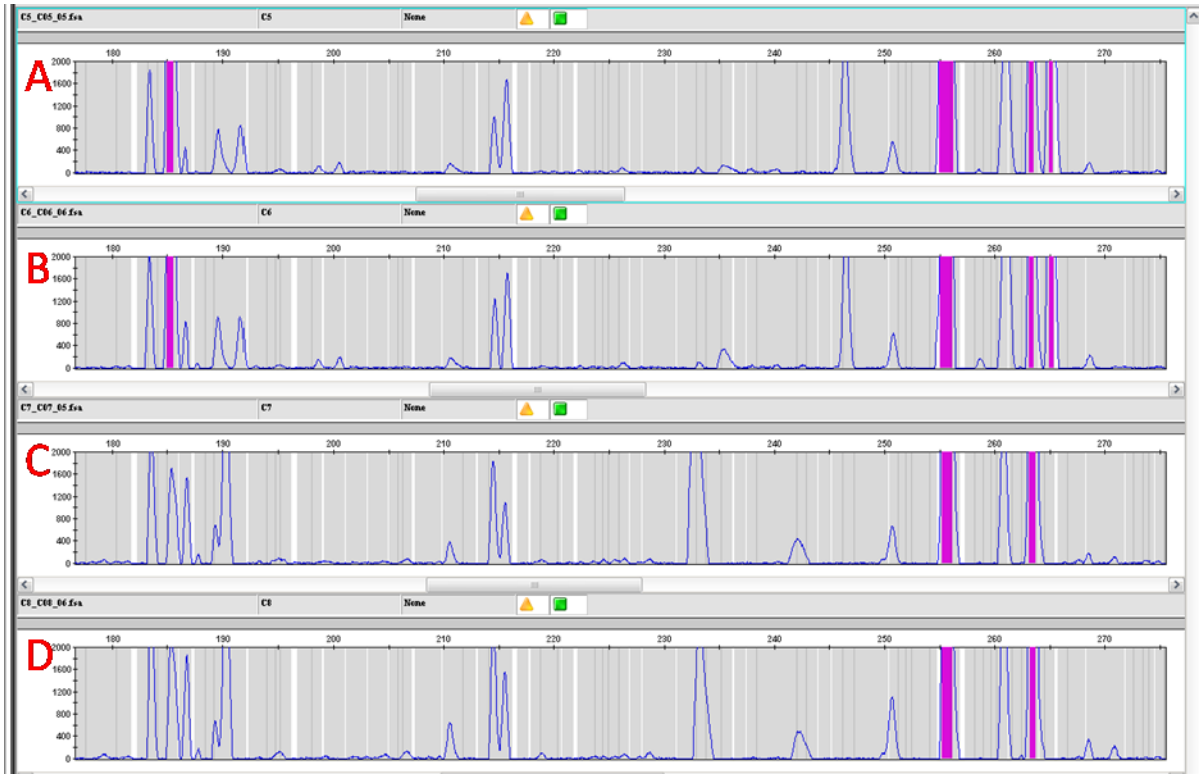
Figure 7. *M. musculus* profile cut with MspI. All panels used primer set Eco-AT/MH-ATT. A and B underwent the same steps up to the select PCR step, which they underwent independently. C and D underwent the same steps up to the select PCR step, which again were run independently. A and B underwent the same reaction as C and D up to the pre-select step in which two aliquots were created and run independently. This created technical replicates for the pre-select, with A and B in one group and C and D in another, and select replicates with A compared to B and C compared to D.



In Figures 7 and 8, we see technical replicates of the pre-select PCR step, the A and B group compared to the C and D group, and the select PCR step, A compared to B and C compared to D. In the MspI enzyme (Fig. 7), we see no differences in any of the profiles. Therefore, we would believe that the method is working extremely well. However when we look at the HpaII enzyme (Fig. 8), we see major differences when we compare the A and B runs to the C and D runs. There are peaks present in both of these groups that are not present in the other. These polymorphisms could cause miscalls and lead to the conclusion that the profile is more highly methylated than it truly is. Differences in the profile only occur at or before the pre-select

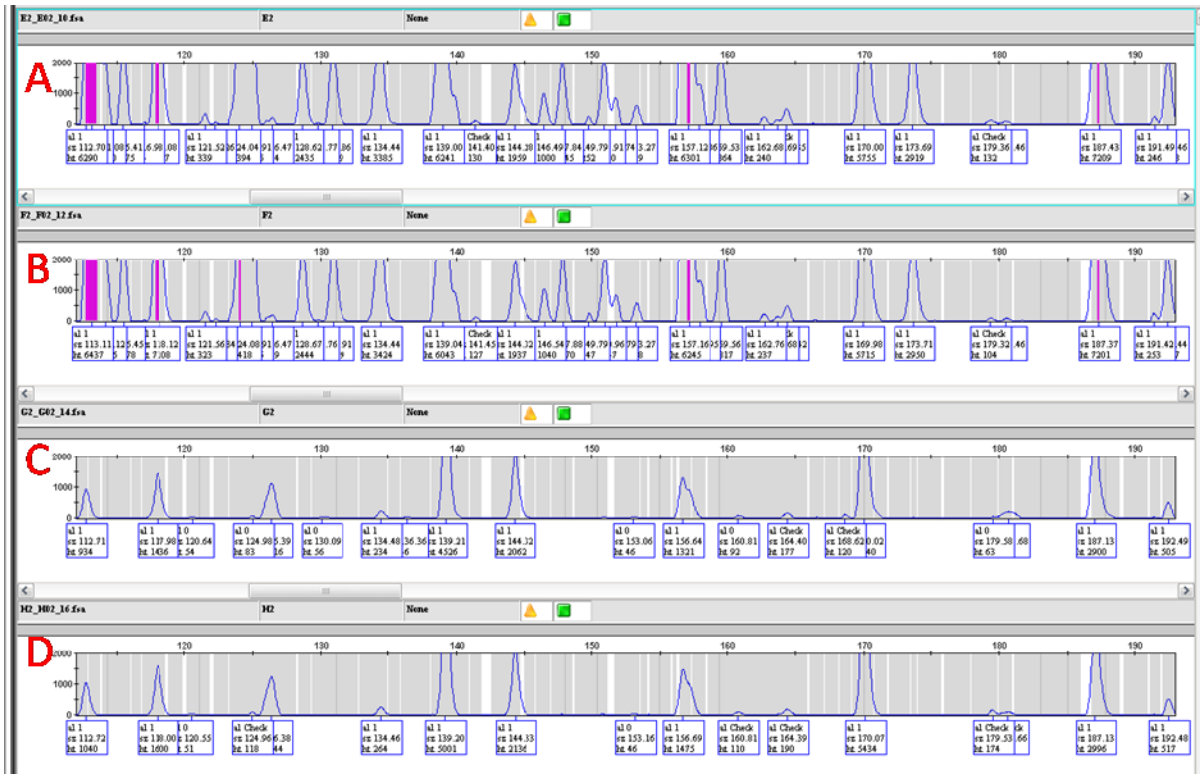
step and the select PCR step shows high reproducibility. Therefore, we can conclude that a step before the select PCR step is causing variation in the HpaII reaction.

Figure 8. *M. musculus* profile cut with HpaII. Eco-AT/MH-ATT. A and B underwent the same steps up to the select PCR step, which they underwent independently. C and D underwent the same steps up to the select PCR step, which again were run independently. This created technical replicates for the pre-select, with A and B in one group and C and D in another, and select with A compared to B and C compared to D.



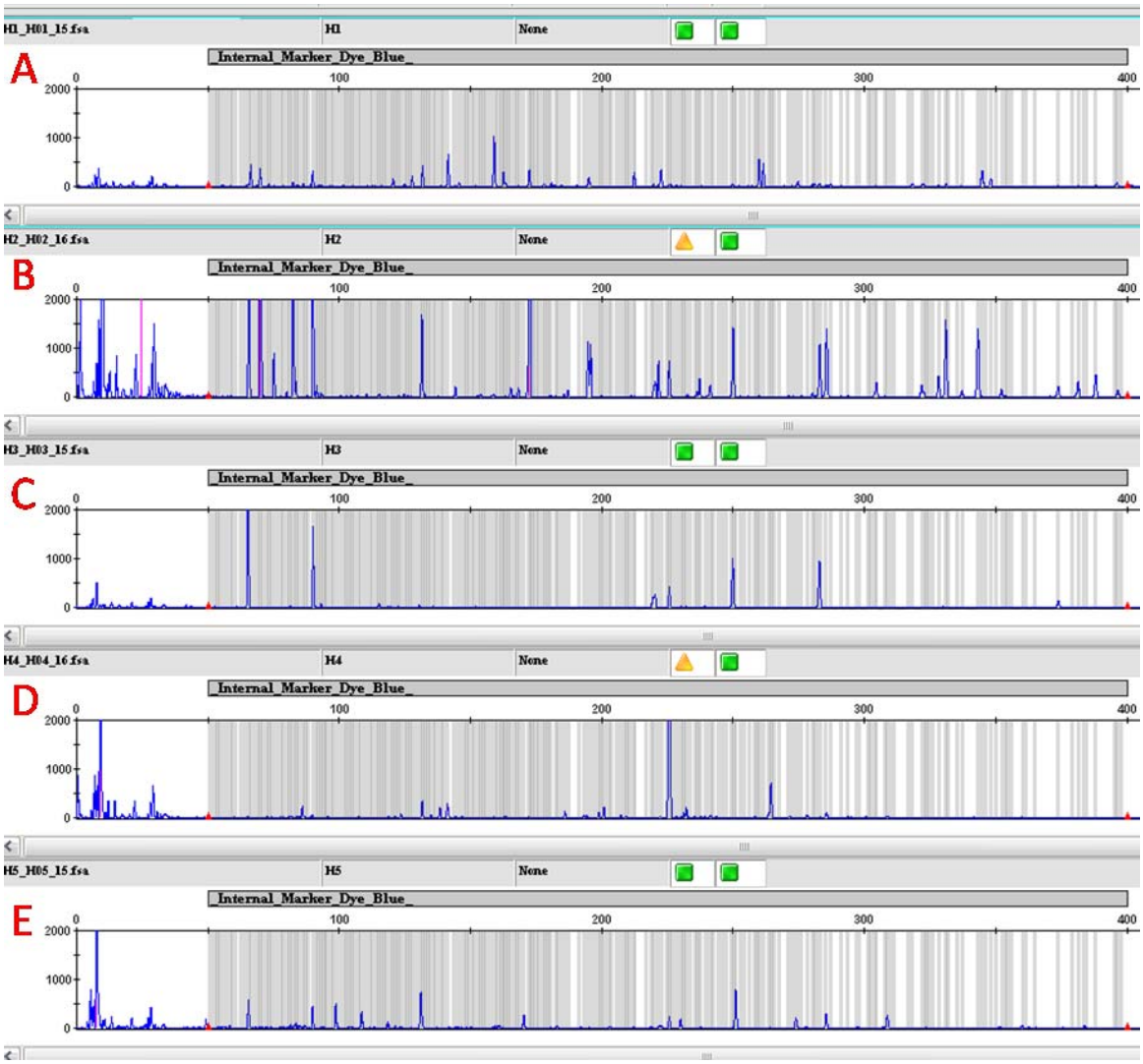
Select-PCR

Figure 9. *H. Sapiens* profile with replicates of select PCR with both HpaII and MspI enzymes under primer set Eco-AG/MH-ATC. A and B are technical replicates of the MspI reaction. C and D are technical replicates of the HpaII reaction. All reactions are run off of the same pre-select PCR sample.



As was shown in Figures 7 and 8, Figure 9 shows very clearly how reproducible profiles from the select PCR step truly are. Fig. 9A and B show almost the exact same profile, as is the case with Fig 9. C and D. In this experiment, we can see polymorphisms between the MspI (A and B) and the HpaII (C and D) reactions. Because of these polymorphisms and the reproducibility of these results, it is clear that the *H. sapiens* genome is methylated to some degree, which is consistent with our expectations.

Figure 10. *D. melanogaster* profile with different number of nucleotide bases used in the select PCR step. All panels used Eco-AG. The panels go in an increasing base number with the MH primer; A-AG, B-ATC, C-ATCA, D-AATAC, E-AATAAC. All steps conducted at the same time and only vary at the select PCR step.

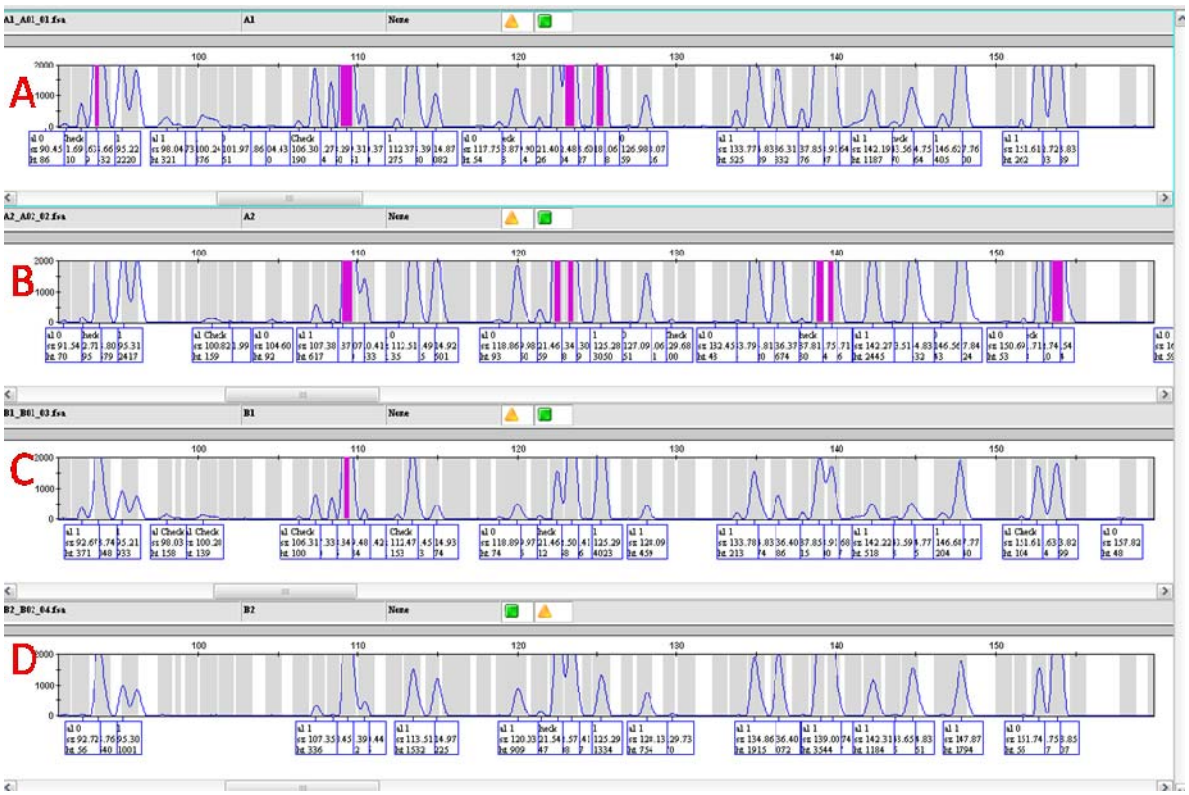


Increasing the number of selective nucleotides, as mentioned in the methods, should decrease the number of peaks in the profile by 1/4 in a nested experiment. In Figure 10, we create a variation of this experiment, with non-nested primers. In this case, we use primers that have two selective nucleotides (Fig. 10A) up to six selective nucleotides (Fig. 10E), in a stepwise fashion. However, it is quite apparent that the step down in peaks present is not occurring. While

we do see a step down between Fig 10B and Fig10C, which are nested, the same step down is not occurring in Fig10E when compared to Fig10D, which are also nested. Therefore, increasing the number of selective nucleotide bases does not appear to be an effective way to control the peak number in AFLP profiles.

Dilution Protocol

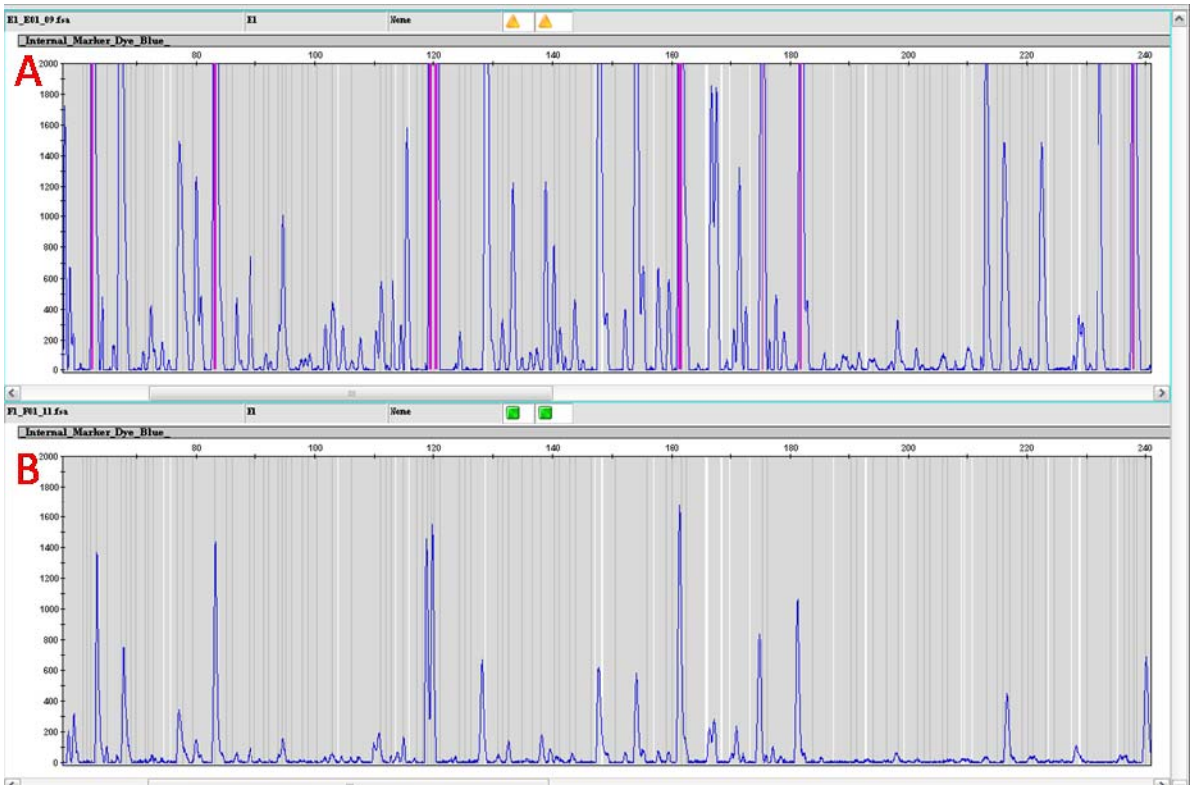
Figure 11. *H. Sapiens* profile cut with the Msp I enzyme, with primer set Eco-AC/MH-ACT. Profiles underwent variable dilution protocol. A and C are the same sample, one diluted, C, and one non-diluted, A. B and D are also the same sample with one diluted, D, and one non-diluted, B. C and D are diluted with 100µL of H_20. Pink lines indicate fragments that exceed the detection limits of the machine.



The next aspect of the AFLP protocol that we experimented with was the dilution protocol after the select PCR step. The results of one of these experiments are shown in Figure 11. Fig. 11A and B, which were not diluted, look much stronger compared to the non-diluted samples, Fig 11C and D, which is as expected. The intensity of the peaks is important because a

sample that is diluted too much could cause miscalls as can be seen in peak present/absence in comparing Fig 11C to D. However, a run with too "hot" of an intensity could cause blowouts in the sequencer, which are shown to a much greater degree in Fig 11A and B. These blowouts could cause a distortion of the read and create extra "false" peaks. Therefore, a compromise needs to be identified to optimize this protocol.

Figure 12. *V. maculifrons* profile cut with the Msp I enzyme, with primer set Eco-AC/MH-ACT. Profiles underwent variable dilution protocol. A is the non-diluted samples. B is the diluted according to the protocol in the methods section. Pink lines are blowouts in the machine.



The need to find a balance is easily seen in Figure 12. The only difference between Fig. 12A and B is the dilution protocol between them. However, this change causes a huge difference in the number and intensity of peaks present. Therefore, we can conclude that the dilution protocol could easily be a source of variation between profiles.

Organism Variation

Figure 13. *P. exclamans* profile with primer set Eco-AG/MH-ACT. The top panel was cut with the MspI enzyme. The bottom was cut with the HpaII enzyme.

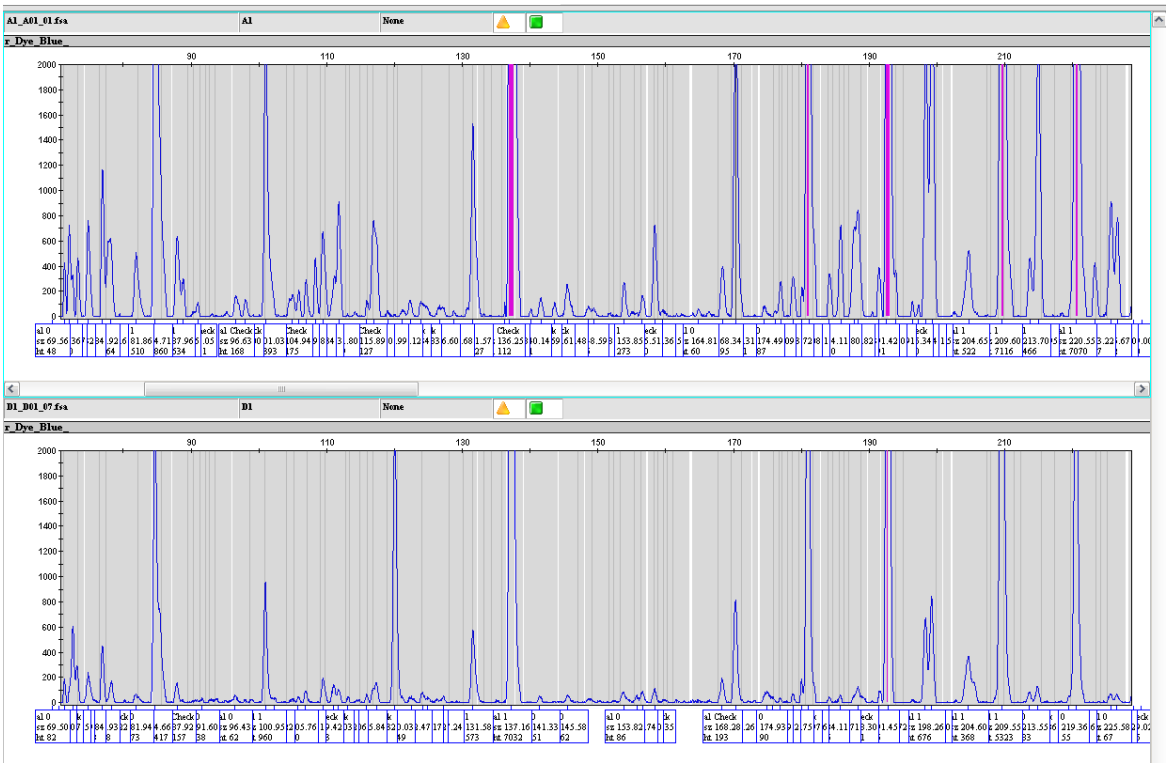


Figure 13 shows the profile of the solitary wasp *P. exclamans*. Of note here is the presence of peaks in the MspI digest, the top panel, that are not present in the HpaII digest, the bottom panel. This presence indicates that this organism is methylated to some degree.

Figure 14. *A. mellifera* profile with primer set Eco-AG/MH-ACT. The top panel was cut with the MspI enzyme. The bottom was cut with the HpaII enzyme.

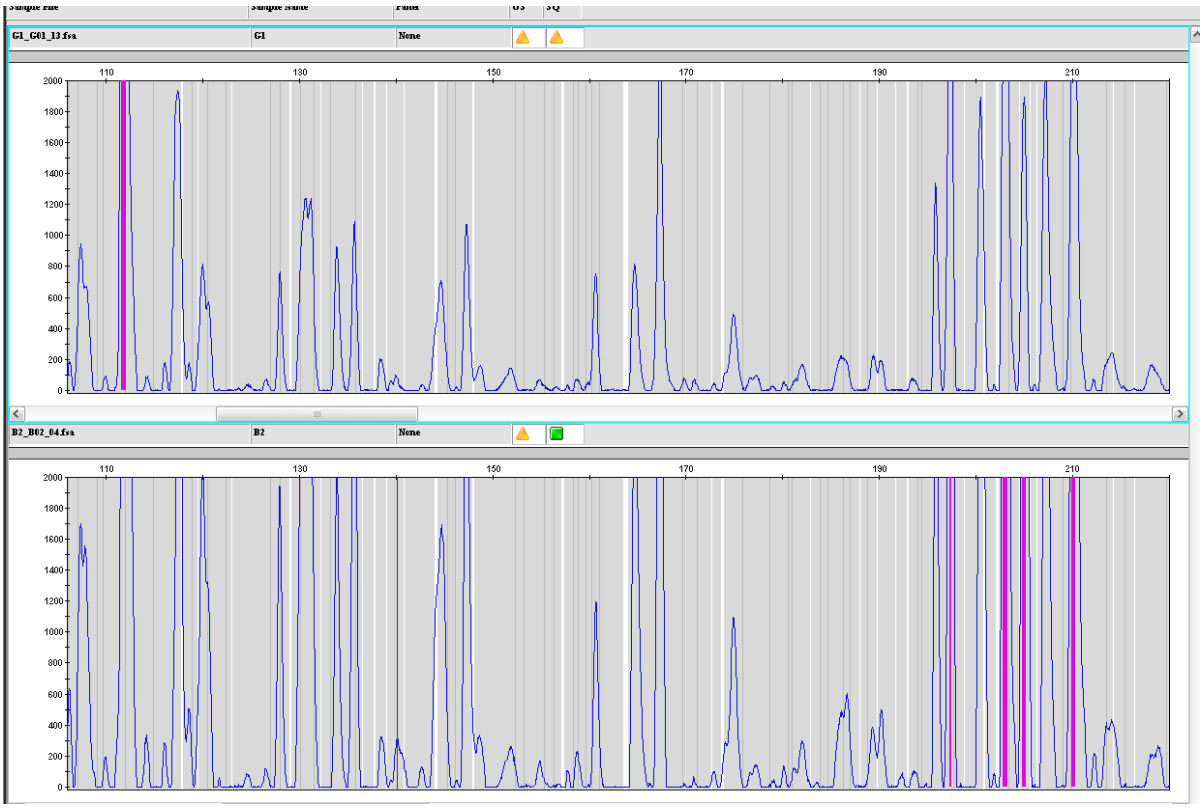


Figure 14 depicts the average run of the honeybee, *A. mellifera*. Contrary to what was seen in Figure 13, we see no differences between the MspI and HpaII profiles, as the profiles are extremely similar to one another. We have consistently seen this result through all of the honeybee AFLP profiles. In addition, we see this same result with *V. maculifrons* queen (Figure 15).

Figure 15. *V. maculifrons* queen profile with primer set Eco-AG/MH-ACT. The top panel was cut with the MspI enzyme. The bottom was cut with the HpaII enzyme.

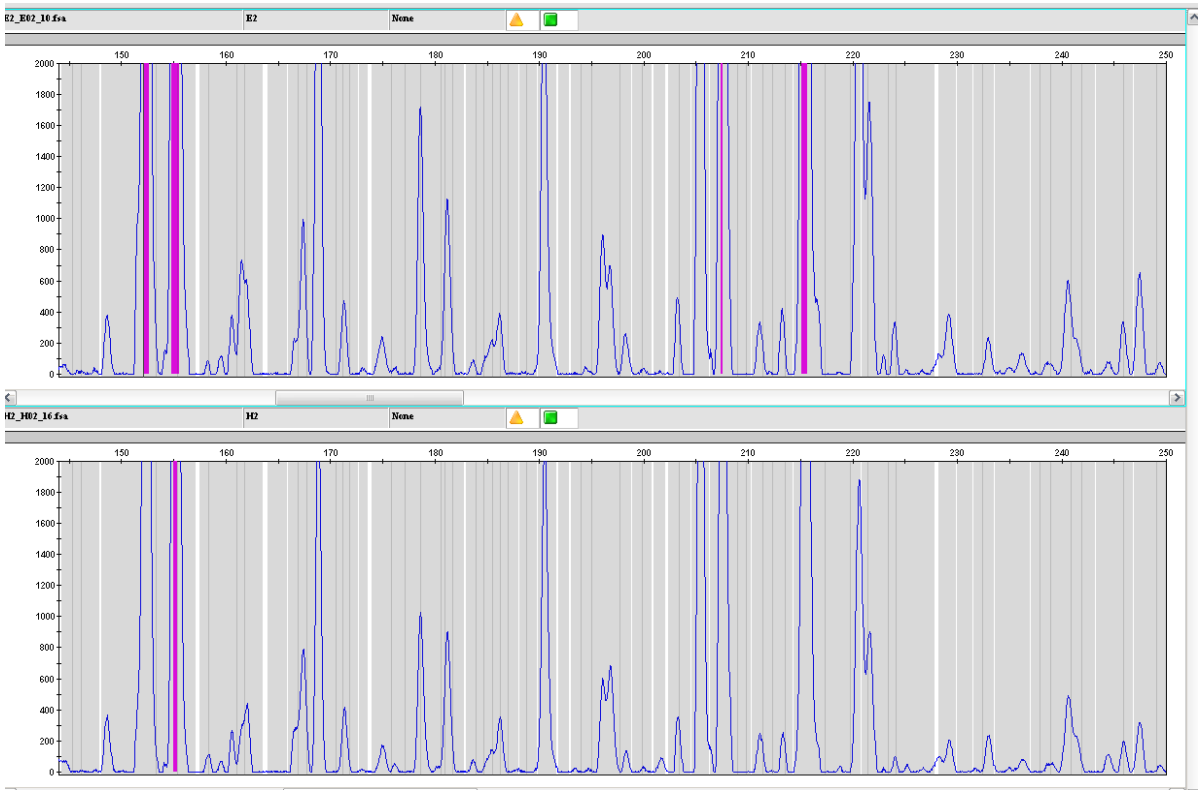
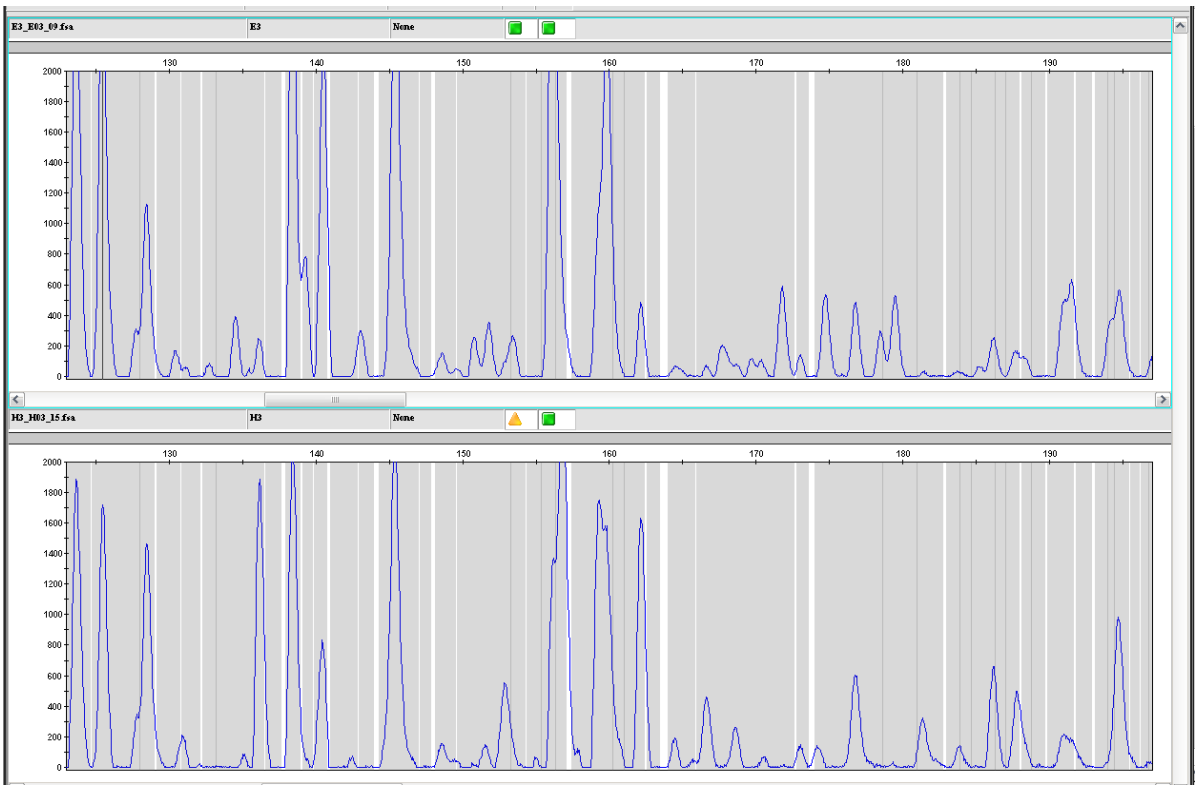


Figure 16. *V. maculifrons* worker profile with primer set Eco-AA/MH-AAT. The top panel was cut with the MspI enzyme. The bottom was cut with the HpaII enzyme.



In the *V. maculifrons* worker and male (Figures 16 and 17, respectively) we see the interesting result of methylation patterns present in the AFLP runs. In the worker we see minimal, yet undeniable, differences between the MspI and HpaII enzymes. In the male, the differences are even more apparent.

Figure 17. *V. maculifrons* male profile with primer set Eco-AA/MH-AAT. The top panel was cut with the MspI enzyme. The bottom was cut with the HpaII enzyme.

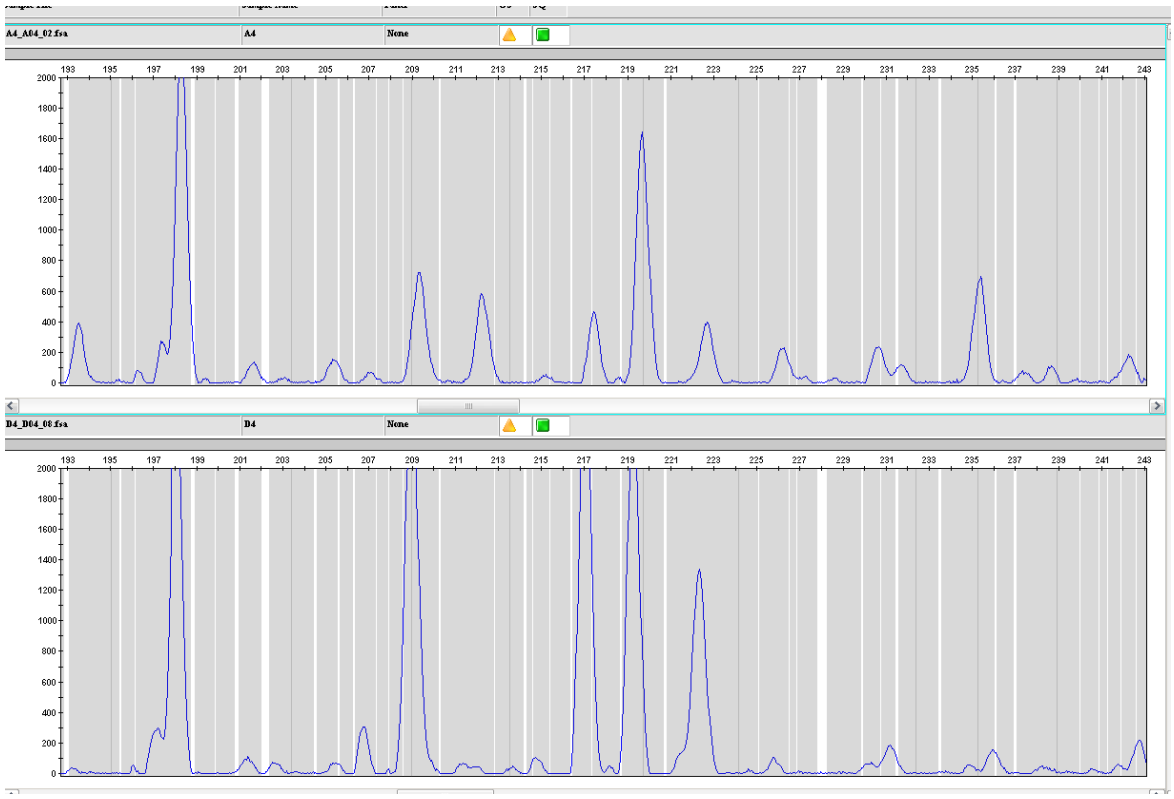
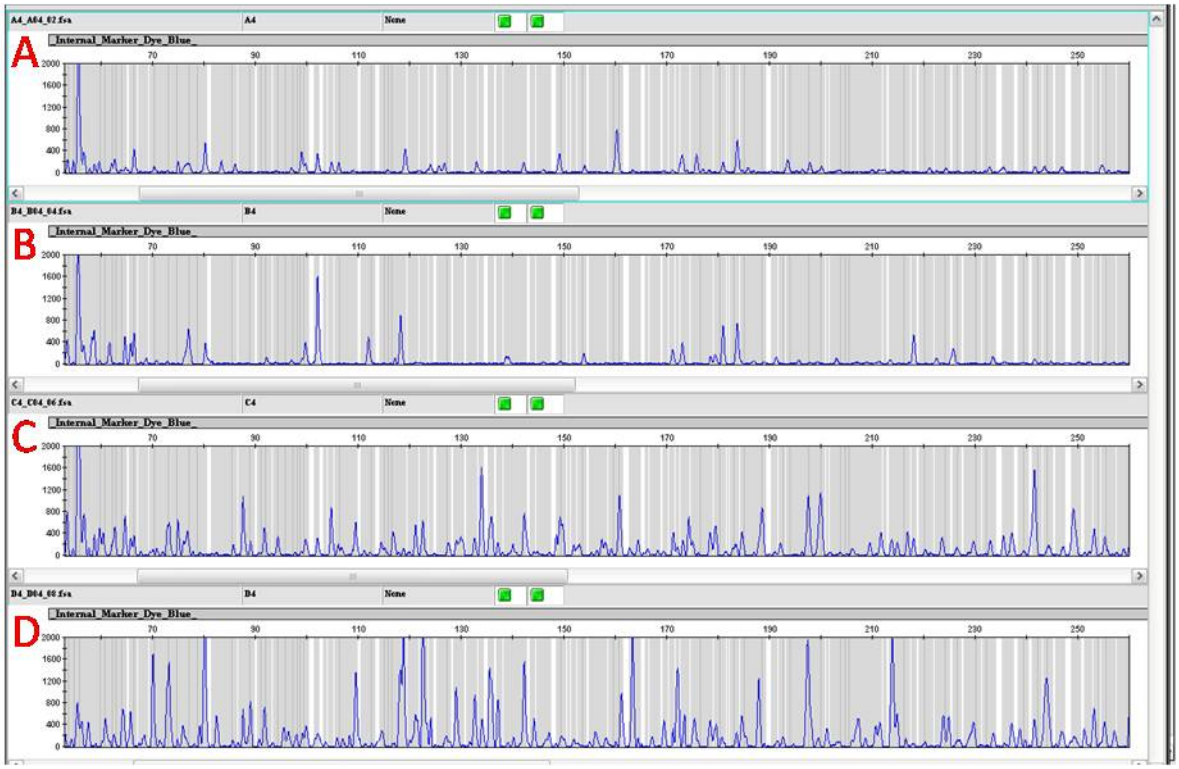
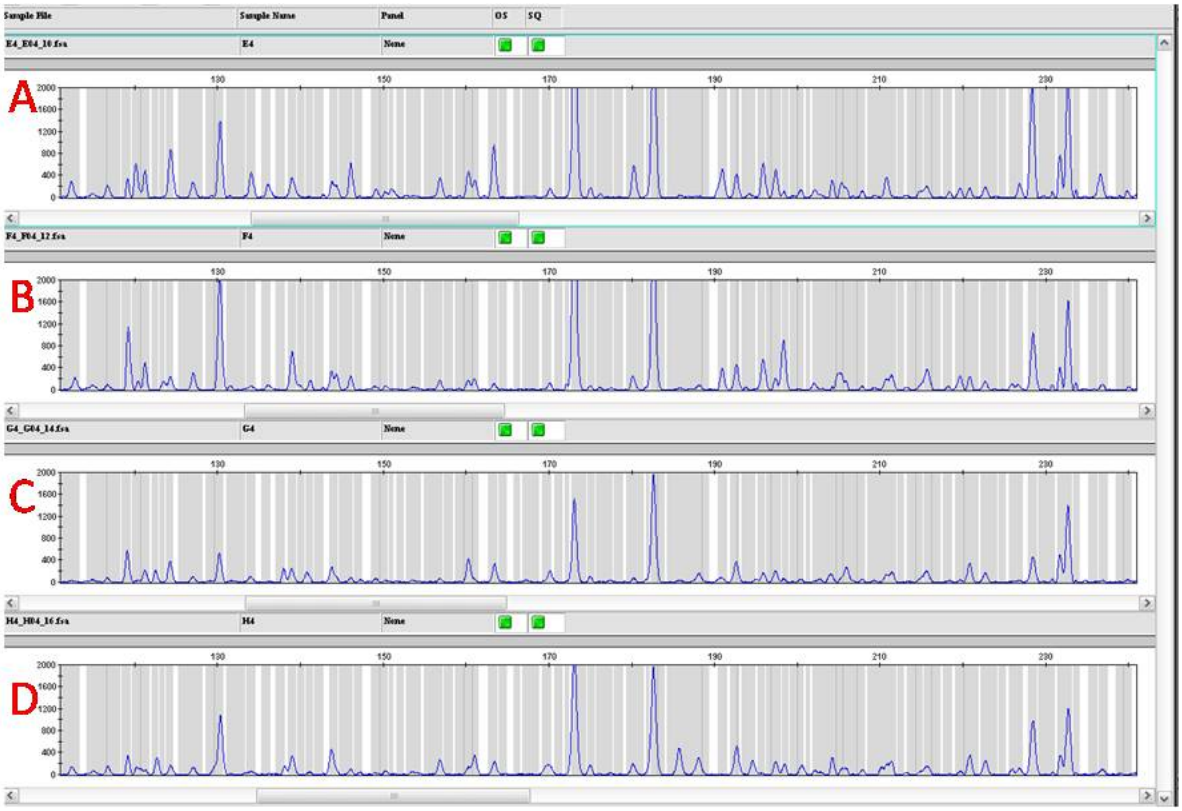


Figure 18. *H. Sapiens* profile with primer set Eco-AT/MH-ACT. A and B are HpaII technical replicates. C and D are MspI technical replicates.



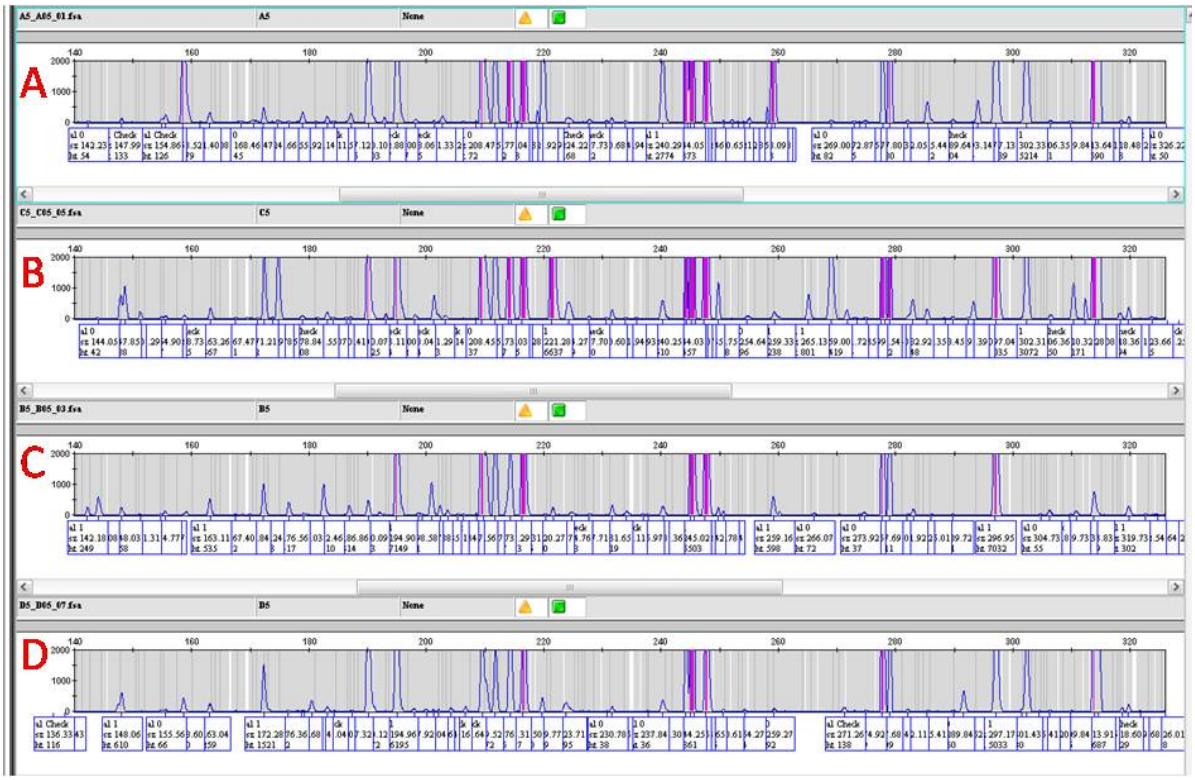
In the next set of experiments, we included technical replicates with our method (Figure 18). The HpaII enzyme digestion replicates, Figs 18A and 18B, have many differences between their profiles. It looks like these differences are due to both peak intensity differences and peak presence and absence. However, we do not see the same result within the MspI replicates and, although we see differences, they are nowhere near the degree of polymorphism within the HpaII replicates.

Figure 19. *V. maculifrons* worker profile with primer set Eco-AT/MH-ACT. A and B are HpaII replicates. C and D are MspI replicates.



Taking another look at the V. maculifrons worker, this time with replicates (Figure 19), we see a story a little different than with the human. The MspI replicates, Figs. 19C and D, look very similar to one another, meaning that they share the same peak profile. The HpaII enzyme replicates, Figs. 19A and B, show some polymorphisms within their profiles but not as many as was seen in Figs. 18A and 18B. Comparing the MspI replicates to the HpaII replicates, we see a high correlation leading to a very low apparent level of methylation when visually comparing it to the differences within enzymes.

Figure 20. *C. Elegans* profile with primer set Eco-AG/MH-AAT. A and B are MspI replicates and C and D are HpaII replicates. These profiles all underwent the same digestion and underwent the rest of the protocol independently.



Because of our problems with controls, we looked at one more model organism with a defined level of methylation, *C. elegans* (Figure 20). Again, we see the same problems as before, peaks present within both replicates of MspI and HpaII (Figs. 20A and B for MspI and C and D for HpaII) and peak presence/absence between the enzymes, which should not be there due to its lack of genomic DNA methylation.

Overall, we are seeing differences in our methylation-sensitive AFLPs due to our selection of organisms that we did not expect. *H sapiens*, *M. musculus*, *D. melanogaster*, and *C. elegans* are all showing differences within their technical replicates. This is a problem that is not occurring in our organisms of interest, *V. maculifrons* and *A. mellifera*.

DISCUSSION

Our experimentation failed to return profiles that we could confidently state provided an accurate portrayal of the known differences in the DNA methylation levels of different species. Of particular concern was our inability to generate profiles devoid of methylation markers in our negative controls. This result has led us to question our protocol and the methodology. The AFLP methodology is known to give rise to some erroneous variation (Laird, 2010). Regardless, few papers devote the time to address these problems or fully discuss their implications, particularly in the realm of methylation-sensitive AFLP (Trybush *et al.*, 2006).

In this study, we took a systematic approach to determining possible sources of error. Through our experiments, we have found many aspects of the methylation-sensitive AFLP that are robust to change and are not sources of variation between profiles. The initial concentration of DNA, ligation reaction temperature, PCR cocktail variation, and the select PCR step do not appear to be significant causes of variation. Surprisingly, the time of digestion in both enzymes do not seem to play much of a role in AFLP polymorphism, which is contrary to results from previous studies (Laird, 2010). Finally, we found that the adding of a selective nucleotide to the select PCR step (Fig. 9) did not decrease the number of profiles but 1/4 as expected, but appeared to have somewhat random effects on the number of DNA fragments. This suggests strong primer-specific differences in the protocol's efficiency. Thus, many of the experimental conditions of methylation-sensitive AFLP appear robust to variation.

We believe our inability to get this technique to work properly arose from several sources of variation, which should be considered by other individuals wishing to use this methodology. For example, the dilution protocol of the select PCR products should be carefully optimized, but

once one finds a balance it should not cause significant variation. The major causes of variation within the technical replicates were the ligation reaction and pre-select PCR (see corresponding results section). The experiment which left out the pre-select PCR step (Fig. 6) did not return any higher degree of replicability in final profiles. Skipping the pre-select PCR step might even cause more variability in profiles due to the resulting low peak intensity.

All of the variations in the technical replicates of the ligation and pre-select PCR reactions occurred mostly, if not exclusively, in the HpaII enzyme. It is unclear why the HpaII enzyme exhibited higher variability or whether that could be associated with its methylation sensitivity. Regardless, more detail should be devoted to optimizing this enzyme in any future studies involving this methodology.

In this study, we observed surprising and unpredicted differences within the profiles of different organisms. If our qualitative results are to be believed, we would infer the greatest to least DNA methylation to occur in the following order: *H. sapiens/M. musculus, P. exclamans, D. melanogaster/C. elegans, A. mellifera/V. maculifrons* worker and male, *V. maculifrons* queen. If we ignore the negative controls, these results are consistent with published results (Kronforst *et al.*, 2008; Kucharski *et al.*, 2008). However, because our negative controls did show substantial unfounded variability, our results must be approached with extreme caution. We believe that our negative control results do not speak to the methods inability to detect reliably methylation levels but that the method did not work when we implemented it.

The methylation-sensitive AFLP methodology also did not return highly reproducible results in our investigation. The *H. sapiens, M. musculus, D. melanogaster*, and *C. elegans* runs all show very poor reproducibility within their technical replicates. Suprisingly, *V. maculifrons* and *A. mellifera* do show high reproducibility within their technical replicates.

Based on our results, we argue that all applications of methylation-sensitive AFLP methodology must include an experimental design with extensive technical replication and appropriate positive and negative controls. Our results suggest that, without these controls, spurious and misleading results can be obtained with methylation-sensitive AFLP. The method has many steps that have the ability to cause variation. If we had not tested technical replicates and control organisms, we could easily have stated that *V. maculifrons* is subject to similar levels of methylation as *A. mellifera* based on our data. Therefore, we conclude that, while the methylation-sensitive AFLP methodology can be an effective tool for detecting the level of global levels of methylation, controls must be incorporated before interpreting results.

ACKNOWLEDGMENTS

This research was supported by the Georgia Institute of Technology and NSF grants IOS-0820645 and DEB-0640690. We would like to thank Danielle Lawson for her comments on an earlier draft, Jennifer L. Kovacs for her help in methodology, and Sarah M. Marriner for her work in experimentation.

REFERENCES

Boks MP, Derks EM, Weisenberger DJ, *et al.* (2009) The Relationship of DNA Methylation with Age, Gender and Genotype in Twins and Healthy Controls. *PLoS ONE* **4**, e6767.

Brena RM, Huang THM, Plass C (2006) Quantitative assessment of DNA methylation: potential applications for disease diagnosis, classification, and prognosis in clinical settings. *Journal of Molecular Medicine-Jmm* **84**, 365-377.

Field LM, Lyko F, Mandrioli M, Prantera G (2004) DNA methylation in insects. *Insect Molecular Biology* **13**, 109-115.

Fraga ME, Esteller M (2002) DNA methylation: A profile of methods and applications. *Biotechniques* **33**, 632-+.

Fuke C, Shimabukuro M, Petronis A, *et al.* (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Annals of Human Genetics* **68**, 196-204.

Han TH, van Eck HJ, De Jeu MJ, Jacobsen E (1999) Optimization of AFLP fingerprinting of organisms with a large-sized genome: a study on Alstroemeria spp. *Theoretical and Applied Genetics* **98**, 465-471.

Johnston JW, Harding K, Bremner DH, *et al.* (2005) HPLC analysis of plant DNA methylation: a study of critical methodological factors. *Plant Physiology and Biochemistry* **43**, 844-853.

Kronforst MR, Gilley DC, Strassmann JE, Queller DC (2008) DNA methylation is widespread across social Hymenoptera. *Current Biology* **18**, R287-R288.

Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**, 1827-1830.

Laird PW (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* **11**, 191-203.

Marhold J, Kramer K, Kremmer E, Lyko F (2004) The Drosophila MBD2/3 protein mediates interactions between the Ml-2 chromatin complex and CpT/A-methylated DNA. *Development* **131**, 6033-6039.

Rozhon W, Baubec T, Mayerhofer J, Scheid OM, Jonak C (2008) Rapid quantification of global DNA methylation by isocratic cation exchange high-performance liquid chromatography. *Analytical Biochemistry* **375**, 354-360.

Schilling E, Rehli M (2007) Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* **90**, 314-323.

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* **9**, 465-476.

Trybush S, Hanley S, Cho KH*, et al.* (2006) Getting the most out of fluorescent amplified fragment length polymorphism. *Canadian Journal of Botany-Revue Canadienne De Botanique* **84**, 1347-1354.

Walsh TK, Brisson JA, Robertson HM*, et al.* (2010) A functional DNA methylation system in the pea aphid, *Acyrthosiphon pisum*. *Insect Molecular Biology* **19**, 215-228.

Wang Y, Jorda M, Jones PL*, et al.* (2006) Functional CpG methylation system in a social insect. *Science* **314**, 645-647.

Werren JH, Richards S, Desjardins CA*, et al.* (2010) Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species. *Science* **327**, 343-348.